ISSN 0317-2473

Research Discussion Paper No. 74

SAMPLING METHODS FOR TELEPHONE SURVEYS

BY

N.M. Lalu

Department of Sociology
University of Alberta
Edmonton, Alberta

January 1991

92-04427/

Sampling Methods for Telephone Surveys.

N. M. Lalu

Abstract

This paper presents three different methods of selecting random telephone numbers for telephone surveys, using relatively simple computer programmes. The design of the computer programmes for this purpose is discussed and sample outputs from three computer programmes, which can be run from an IBM PC, are given.

Introduction

Many researchers with limited budgets use telephone interviews to collect data. Compared to face-to-face or mail surveys, telephone surveys are fast and cheaper. The principal objection to telephone surveys has been the bias introduced by the non-coverage of households without telephones (Sykes and Collins, 1987). In North America more than 90 percent of the households can now be reached by telephone, making the bias due to incompleteness of the sampling frame almost negligible (Sudman, 1973; Tuchfarber and Klecka, 1976). Another problem with sampling for telephone surveys, especially in large cities, is the increase of unlisted numbers. A large percentage of households with unlisted numbers makes the telephone directory a defective sampling frame for the selection of sample telephone numbers.

There are two approaches to sampling for telephone surveys. In the first method, the listed telephone numbers of an area are used as the sampling frame. An alternate approach is the random selection of digits to form sample telephone numbers. There are many variations and combinations of the two approaches. Random digit dialing has been advocated by many researchers to overcome the problem of unlisted telephone numbers (Glasser and Metzger, 1972). Random digit dialing overcomes the problem of unlisted telephone numbers, but may generate many non-working numbers. Use of public directories largely eliminates non-working numbers but excludes unlisted numbers. The question of which method is most useful in a particular case depends mainly on the proportion of unlisted numbers. This paper discusses three methods of selecting sample telephone numbers and three computer programmes to simplify this process.

Telephone directory method

This is the quickest and easiest method of selecting a random sample of telephone numbers for a survey. Most of the selected numbers will be usable because they are obtained from the published directory. The direct application of simple random sampling technique will involve the counting of the telephone numbers and associating a sequence number with each one, a very time consuming process. One may simplify the selection process by applying the principles of cluster sampling. The pages of a telephone directory may be considered to be the clusters and random pages can be selected. In general, cluster sampling leads to higher sampling error than simple random sampling (Kish, 1965). In the following section a modified method of selecting a simple random sample of telephone numbers from a directory is discussed.

Usually telephone directories have a fixed number of columns in each page and a limited number of lines per column. This feature can be effectively utilized to get a sample of telephone numbers. First, determine the number of pages, the number of columns and the maximum number of lines per column. Let these numbers be P, C and L respectively. Then a vector (with three elements) of random integers is selected so that the first element lies between one and P, the second element is between one and C, and the third one falls between one and L. These three integers represent a random page, column and line in the telephone directory and thus identify a random telephone number. The process may be repeated as many times as is necessary to obtain the desired sample size, deleting any duplication in the selected triplets.

A computer programme can simplify the process of obtaining random page, column, and line numbers. SELPAGE1 is a programme that is designed to carry out the selection of a random sample of telephone numbers in this manner. This programme deletes any duplicate selection and prints out the page, column, and line numbers. A sample out-put from this programme is shown below

Write a random number. It should be an odd number
less than 32000
321
How many pages are there?
300
How many columns in each page?
4
How many lines in each column?
80

Write the sample size, it should be less than 10001
5
The random numbers may be written to a file, on the screen or to a printer
1. file
2. printer
3. screen
Write 1, 2 or 3
2
Selected random numbers are:-

Sample No.	Page	Column	Line
1	243	2	61
2	24	1	1
3	213	3	16
4	157	4	35
5	191	2	26

This selection of sample pages assumes that the page numbers start at one and end at the number given for the number of pages. It may happen that the listing of the telephone numbers begin at a different page than the first one in the directory. A second modified programme (SELPAGE2), which takes this factor into account, is available. The print out from this programme is similar to the one given above, except that a constant number, representing the number of blank or unused pages in the beginning of the directory, is added to the random page numbers. The user is prompted to give the page number in the directory where the listing begins.

In a telephone directory, some names and addresses may be printed using large fonts and consequently may be spread over two or three lines. Care should be taken to associate one telephone number with a single line while selecting a sample. A number that is printed over two lines should be included in the sample only if the first line is selected. This will give equal chance for all numbers to be included in the sample. If a selected line number is blank or has the number of a business, one should take a new page, column, and line number as a replacement (never choose the next line as a substitute).

The use of the telephone directory is efficient in the sense of providing a sample with few non-working numbers. Yet the sample can be deficient if the proportion of unlisted numbers is large. In large cities and suburbs as many as 40% of all telephones

may be unlisted (Rossi et. al., 1983, p. 178). In such places the telephone directory method is unsatisfactory.

Random digit dialing

There are several variations of random digit dialing (RDD). In the simplest form RDD ignores the telephone directory completely, using the list of exchanges represented by the first three digits of telephone numbers as the first stage of selection. The exchanges are selected with replacement (which means that a telephone exchange can be selected more than once) and a four digit random number between 0000 and 9999 is appended to each of them. Thus all telephone numbers, whether unlisted or not, have an equal chance of being selected. At the same time, this method will require the screening of a large number of telephone numbers to get a sample of working telephone numbers. The Survey Research Center of the University of Michigan obtained only 22 percent of randomly selected numbers as working numbers (Groves and Kahn, 1979)

The hit rate of working numbers can be improved by the use of a combination of telephone directory and random digit generation methods (Sudman, 1973). The basic idea is to replace the last part (two or three digits) of the listed telephone numbers by random digits. The annual Alberta Survey conducted by the Population Research Laboratory applies this method to select telephone numbers. The procedure used in the Alberta Survey will be described in detail later.

Two different methods of generating random telephone numbers for surveys are presented here. The procedure for computerized selection and sample output from two computer programmes designed for this are given in the following paragraphs.

Random telephone numbers using telephone exchange prefixes

A list of the first three digits of telephone numbers denoting the exchanges in a geographical target area can be usually found in the first several pages of the local directory. Once this list has been constructed, the procedure for generating the numbers is straightforward. First, one exchange is selected at random from the list of exchanges. Then a four digit random number should be produced to complete the telephone number. The required number of telephone numbers can be obtained by

repeating this sequence. The exchanges should be randomly selected anew for each number, and duplicates removed from the sample.

The computer programme RDD simplifies these three steps and generates as many random telephone numbers as necessary. The user has to give a list of three digit numbers representing the exchanges and a random number as a seed to generate the random telephone numbers. The prefixes representing the exchanges can be stored in a computer file or typed in from the key board. The output may be directed to another file or to a printer. A sample printout from the programme RDD is shown below. Here five random telephone numbers are produced from a list of two telephone exchanges.

This programme generates a set of telephone numbers
for random digit dialing using a number of exchanges.
The user should provide the exchange prefixes and
a random number as a seed.
Write the number of Exchanges.
2
The three digit prefixes of the exchanges may be read
from a file or from the console.
1 from console
2 from a file
Type 1 or 2
2
Write the three digit prefixes of the two exchanges
433 437
Write an odd number as a random start
3175
Write the number of telephone numbers required
5
The telephone numbers may written to a
1 file
2 printer or
3 screen
write 1, 2 or 3
3
The Exchanges are:
433 437
The Telephone numbers are:
1. 433-8859
2. 437-0434
3. 437-9032
4. 433-5091
5. 433-8363

Random telephone numbers using banks from directories

If sample telephone numbers are exclusively drawn from listed telephone numbers, all unlisted numbers will be ineligible for selection. If the sample is drawn by the RDD technique, all unlisted as well as listed numbers have equal chance of being included in the sample. However, a large proportion of the sample numbers may not be working numbers and several of them may be assigned to businesses. The need for screening large lists of telephone numbers can be reduced by a combination of telephone directories and random assignment of digits. In this method, the last two or three digits of telephone numbers in a directory are replaced by random digits. This procedure tries to take advantage of the practice of telephone companies of assigning telephone numbers to customers in blocks of consecutive numbers.

The annual Alberta Survey conducted by the Population Research Laboratory of the University of Alberta uses the method of assigning two random digits to the last part of telephone numbers published in the relevant directories. The first step is to identify all five-digit banks from the telephone directories. In some areas special directories (such as a street index directory) may be available with all telephone numbers listed sequentially in ascending order. It is a simple matter to make a list of all five-digit banks from this type of directory. If such a sequential list is not available, the telephone numbers should be sorted, the last two digits should be stripped off, and all duplicates removed. Ideally one would sort all numbers in the directory. But this may be too expensive if the telephone numbers are to be entered manually into a computer file. For the Alberta Survey, it has been found that a sample of 20% of the telephone numbers will identify almost all of the five-digit banks.

The second step in selecting a sample of telephone numbers is to choose a simple random sample (with replacement) of five-digit banks. Then a two-digit random number between 00 and 99 is appended to each of the banks selected to get a sample of seven-digit telephone numbers. Any duplicate numbers should be removed.

Two computer programmes have been designed to generate random telephone numbers in this manner. The first programme CRERAF requires a sequential file of five-digit banks as an input file. This programme creates a random access file of these banks of digits. This random access file is later read by the second programme, DRSAMP, which draws the required number of random telephone numbers. An

example of the two steps involved is shown below. Here we have 35 five-digit banks. A sample of 20 random telephone numbers is drawn.

List of five digit banks.

42000
42001
42002
42003
42004
.....
42102
42103

These numbers are saved in a file named BANKS. Now one runs the programme CRERAF which creates a random access file (here it is given the name EXAMP.RAF). The following lines show the output from this programme.

Sequential file with bank? banks
Random access file to be created? examp.raf
Number of banks in the sequential file 35

The Random access file created in the above step is used by the programme DRSAMP to draw random telephone numbers. This programme asks for the file names, the number of banks, the sample size and a random number as a starting point. The following lines show the screen image while running this programme.

Name of the random access file? examp.raf
Name of the output file? samp.tel
How many banks in this file? 35
Sample size? 20
A random number? 3179

Now the file SAMP.TEL contains a sample of 20 random telephone numbers. The first 5 lines of this files are given below.

1 420-0927
2 420-6945
3 420-0549
4 420-6350
5 420-0502

Copies of the computer programmes described here may be obtained from the author at the Population Research Laboratory.

References

Glasser, G. J and G. D. Metzger (1972) "Random Digit dialing as a Method of Telephone Sampling". Journal of Marketing Research 9: 59 - 64

Groves, M. Robert and Robert L. Kahn (1979) "Surveys by Telephone." New York: Academic press.

Kish, L. (1965) "Survey Sampling". New York: Wiley

Rossi, H. Peter, James D. Wright and Andy B. Anderson (1983) "Handbook of Survey Research". New York: Academic Press.

Sudman, S. (1973) "The Uses of Telephone Directories for Survey Sampling". Journal of Marketing Research 10: 204 - 207

Sykes, W. M., and Collins, M. (1987) "Comparing Telephone and Face-to-Face Interviewing in the United Kingdom". Survey Methodology 13: 15 - 28

Tuchfarber, Alfred J and William R. Klecka. (1976) "Random Digit Dialing: Lowering the Cost of Victimization Surveys". Police Foundation: University of Cincinnati.

RESEARCH DISCUSSION PAPER SERIES

59. A Study of the Contemporary Population Migration in the People's Republic of China.
Yihua Yuan. May 1989

60. Might Suicide Underreporting Affect Hypothesis Test? The Case of Suicide and Divorce.
Mark Speechley and Frank Trovato. July 1989.

61. Demographic Concerns in Belles-Lettres: Insights From the Writings of Günter Grass and Margaret Atwood.
Karol Krótki and Guy Germain. January 1990.

62. Some Lacunae in Demographic Data Collection in South Asia with Special Reference to India.
Parameswara Krishnan. January 1990.

63. The Emergence of Multiethnicities in 1981 and 1986: Their Sociocultural Significance.
Karol Krótki and Dave Odynak. February 1990.

64. Dynamics of Internal Migration in Bangladesh.
A.K.M. Nurun Nabi. April 1990.

65. Why Are Canadians Dying Out?
Karol Krótki. April 1990.

66. A Theoretical Model for the Study of Migration and Urbanisation
David Achanfuo-Yeboah. May 1990.

67. Political Demography of the Emergence of Bangladesh
A.K.M. Nurun Nabi and P. Krishnan. May 1990.

68. Dependency Ratios: Conventional and Real.
Parameswara Krishnan and Manoshi Das. June 1990.

69. Log-Linear Modelling and Simultaneous Standardization of Rates.
Frank Trovato. November 1990.

70. Unemployment and Occupational Aspirations: A Panel Study of High School Graduates.
Susan Empson-Warner and Harvey Krahn. November 1990.

71. An Analysis of Native Mortality in Canada.
Andrea Werner-Leonard and Frank Trovato. November 1990.

72. Ethnic Variations in Fertility: Microeconomic and Minority Group Status Effects.
Tina W.L Chui and Frank Trovato. December 1990.

73. Quantifying Semi-Structured Interviews.
Harvey Krahn. December 1990.